

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:16:24 ; Search time 91.75 Seconds

(without alignments)
15.839 Million cell updates/sec

Title: US-09-331-631a-1_COPY_29_73

Perfect score: 252
Sequence: 1 SEFDROETEECKRQCMQLET.....RCVSQCKRFEEDIMSKYD 45

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	77.5	30.8	588	1 VCLB_GOSHI	P09801 gossypium h
2	72	28.6	605	1 VCLA_GOSHI	P09789 gossypium h
3	67.5	26.8	47	1 AGRP_LURCY	P36368 lufta cylin
4	67	26.6	1170	1 TSP1_HUMAN	P07996 homo sapien
5	67	26.6	1170	1 TSP1_MOUSE	P35441 mus musculu
6	65	25.8	1173	1 TSP1_XENLA	P35448 xenopus lae
7	63	25.0	524	1 SBP_SOYBN	O04672 glycine max
8	57	22.6	724	1 ATIL_VACCV	P24759 vaccinia vi
9	57	22.6	726	1 ATIL_CAMPC	O05482 camelpox vi
10	57	22.6	1284	1 PPI_COMPX	P16602 compox viru
11	55	21.8	305	1 PPP6_RAP	O64620 rattus norv
12	55	21.8	795	1 ENPL_CHICK	P08110 gallus gall
13	54.5	21.6	244	1 YL53_CAEEL	P34433 caenorhabdi
14	54.5	21.6	1021	1 YL88_CAEEL	P46582 caenorhabdi
15	54	21.4	291	1 CGD1_BRARE	O90459 brachydanio
16	54	21.4	643	1 R060_CAEEL	O27274 caenorhabdi
17	53.5	21.2	301	1 G197_HUMAN	O14457 homo sapien
18	53.5	21.2	1170	1 TSP2_BOVIN	O95116 bos taurus
19	53	21.0	702	1 ATIL_VARY	P34011 variola vir
20	52.5	20.8	91	1 UCRH_HUMAN	P07919 homo sapien
21	52.5	20.8	1172	1 TSP2_HUMAN	P35442 homo sapien
22	52	20.6	33	1 MBP1_MAIZE	P28794 zea mays (m
23	52	20.6	37	1 CG2S_LUPAN	P09930 lupinus ang
24	52	20.6	82	1 C2_OXYNO	P05526 oxytricha n
25	52	20.6	716	1 ENPL_RABIT	O18750 oryctolagus
26	52	20.6	802	1 ENPL_MOUSE	P08113 mus musculu
27	52	20.6	803	1 ENPL_HUMAN	P14458 homo sapien
28	52	20.6	804	1 ENPL_CANFA	P14148 canis fami
29	52	20.6	1188	1 TIR_ECOLI	P08956 escherichia
30	51.5	20.4	106	1 COLA_HORSE	P02704 equus cabal
31	51.5	20.4	108	1 COLB_HORSE	P02705 equus cabal
32	51.5	20.4	147	1 LAPF_HAEOP	O01747 haemeteria
33	51.5	20.4	298	1 HIST_CANAL	P46386 candida alb

34	51.5	20.4	525	1 Y107_METJA	O57571 methanococ
35	51.5	20.4	563	1 MUCS_BOVIN	P98031 bos taurus
36	51.5	20.4	860	1 UL52_HSV6U	P52467 herpes simp
37	51.5	20.4	860	1 UL52_HSV6Z	P52467 herpes simp
38	51.5	20.4	1557	1 DVAL_DICVI	O24702 dictyocaulu
39	51	20.2	418	1 CDL5_HUMAN	O14004 homo sapien
40	51	20.2	741	1 FIBA_CHICK	P14448 gallus gall
41	51	20.2	881	1 ARP8_YEAST	O12386 saccharomyc
42	51	20.2	893	1 YM92_CAEEL	P34531 caenorhabdi
43	51	20.2	1231	1 CEAH_HUMAN	P08603 homo sapien
44	51	20.2	1370	1 Z261_HUMAN	O14202 homo sapien
45	50.5	20.0	264	1 RPOD_METTH	O26144 methanobact

ALIGNMENTS

```

RESULT 1
VCLB_GOSHI          STANDARD;          PRT;          588 AA.
ID VCLB_GOSHI
AC P09801;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE VICILIN C72 PRECURSOR (ALPHA-GLOBULIN B).
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; gossypium.
RN [1]
RP SEQUENCE FROM N.A.
RA Chlan C.A., Pyle J.B., Legocki A.B., Dure L. III;
RT "Developmental biochemistry of cottonseed embryogenesis and
RT germination. XVIII. cDNA and amino acid sequences of the members of
RT the storage protein families."
RL Plant Mol. Biol. 7:475-489(1986).
CC -I- FUNCTION: SEED STORAGE PROTEIN.
CC -I- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC BODIES.
CC -I- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASOLIN, VICILIN,
CC CONVICTILIN, CONGLYCININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M16891; AAA33071.1; -.
DR PIR; A30838; FMCNAB.
DR HSSP; P50477; ICAX.
DR INTERPRO; IPR001113; -.
DR PFAM; PF00546; Seedstore_7s; 1.
KW Seed storage protein; Signal.
FT SIGNAL 1 25
FT CHAIN 26 588
FT SEQUENCE 588 AA; 69729 MW; 63E699B29A8ADEB CRC64;

```

Query Match 30.8%; Score 77.5; DB 1; Length 588;

Best Local Similarity 40.5%; Pred. No. 0.059;

Matches 15; Conservative 9; Mismatches 12; Indels 1; Gaps 1;

QY 2 EFDRQETEECKRQCMQLETSGOMRCVSCQDKRFEED 38

DB 81 EDPQRYEECKRQEE-EPQPCQCRCKRFEED 116

RESULT 2

VCLA_GOSHI STANDARD; PRT; 605 AA.

```
AC P09799; (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, last sequence update)
DT 15-JUL-1999 (Rel. 38, last annotation update)
DE VICLIN GC72-A PRECURSOR (ALPHA-GLOBULIN A).
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; Gossypium.
RN [1]
RP SEQUENCE FROM N.A.
RA Chan C.A., Borroto K., Kamalay J.A., Dure L. III;
RT "Developmental biochemistry of cottonseed embryogenesis and
RT germination. XIX. Sequences and genomic organization of the alpha
RT globulin (Vicilin) genes of cottonseed.";
RL Plant Mol. Biol. 9:533-546(1987).
CC -!- FUNCTION: SEED STORAGE PROTEIN.
CC -!- SUBCELLULAR LOCATION: CYTOSOL/MEMBRANE-BOUND VACUOLAR PROTEIN
CC BOBIES.
CC -!- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICILIN, CONGLYCININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M19378; AAA33069.1; -.
DR PIR: S06398; S06398.
DR HSSP: P50477; ICAX.
DR INTERPRO: IPR001113; -.
DR Pfam: PF00546; Seedstore_7s; 1.
KM Seed storage protein: Signal.
FT SIGNAL 1 23
FT CHAIN 24 605 VICILIN GC72-A.
FT FT 605 AA; G9DB937JC976953B CRC64;
SQ SEQUENCE 605 AA; 71049 MW; C9DB937JC976953B CRC64;

Query Match 28.6%; Score 72; DB 1; Length 605;
Best Local Similarity 42.5%; Pred. No. 0.26;
Matches 17; Conservative 9; Mismatches 10; Indels 4; Gaps 3;

OY 1 SEFD-RQEYEBCCKRCQMLETSGCMRRCVSOCDKRFEEEDI 39
   || | :| |:| :||| ||| || :| :| | :
Db 33 SEDDPQORYEDCDKRKC-OLETRGOTEG--DKCEDRSFQL 69

RESULT 3
ACRP_LUCFCY STANDARD: PRT; 47 AA:
ID_AGRP_LUCFCY P56568;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 6.5 kDa AGCININE/GUTAMATE-RICH POLYPEPTIDE (6.5k-AGRP).
OS Luffa cylindrica (Smooth loofah) (Sponge gourd).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Cucurbitales; Cucurbitaceae; Luffa.
RN [1]
RP SEQUENCE.
RP TISSUE=SEED;
RX MEDLINE=97357433; PubMed=9214759;
RA Kimura M., Park S.S., Sakai R., Yamasaki N., Funatsu G.;
RT "Primary structure of 6.5k-arginine/glutamate-rich polypeptide from
RT the seeds of sponge gourd (Luffa cylindrica)".;
RL Biosci. Biotechnol. Blochm. 61:984-988(1997).
CC -!- FUNCTION: STORAGE PROTEIN WHICH PROVIDES NITROGEN AND CARBON
CC RESERVES DURING GERMINATION AND SEEDLING GROWTH.
CC -!- MASS SPECTROMETRY: MM=5693.39; METHOD=WALDI.
```

```

CC      -1- SIMILARITY: SOME, TO 7S SEED STORAGE PROTEINS.
KM      Seed storage protein.
FT      DISULFID 12 33
FT      DISULFID 16 29
SQ      SEQUENCE 47 AA; 5698 MW; 588B0EC82273AC05 CRC64;

Query Match      26.8%; Score 67.5; DB 1; Length 47;
Best local similarity 41.2%; Pred. No. 0.081;
Matches 14; Conservative 4; Mismatches 15; Indels 1; Caps

QY      5 ROEYEECKRCQCMOLETS-GQMRRCVSOCDKRFEE 37
          | | | | : | | | : | | | | | | |
Db       6 RTEYECRCRCVAEHGVERRCQYCEKRLRE 39

RESULT 4
TSPL_HUMAN TSPL_HUMAN STANDARD: PRT: 1170 AA.
AC          P07996;
DT          01-AUG-1988 (Rel. 08, Created)
DT          01-AUG-1988 (Rel. 08, Last sequence update)
DT          01-OCT-1996 (Rel. 34, Last annotation update)
DE          THROMBOSPONDIN-1 PRECURSOR.
GN          THBS1 OR TSPL OR TSP.
OS          Homo sapiens (Human).
OC          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
NC          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
[1]
RP        SEQUENCE FROM N.A.
RC        TISSUE-ENDOTHELIAL CELLS;
RX        MEDLINE-87057617; PubMed-2430973;
RA        Lawler J., Hynes R.O.;
RT        "The structure of human thrombospondin, an adhesive glycoprotein with
RT        multiple calcium-binding sites and homologues with several different
RT        proteins.";
RL        J. Cell Biol. 103:1635-1648(1986).
[2]
RN        SEQUENCE FROM N.A.
RP        MEDLINE-89139590; PubMed-2918029;
RX        Hennessy S.W., Frazier B.A., Kim D.D., Deckwerth T.L.,
RA        Baumgartel D.M., Rotwein P., Frazier W.A.;
RT        "Complete thrombospondin mRNA sequence includes potential regulatory
RT        sites in the 3' untranslated region.";
RL        J. Cell Biol. 108:729-736(1989).
[3]
RN        SEQUENCE OF 1-397 FROM N.A.
RP        MEDLINE-87157592; PubMed-3030396;
RX        Kobayashi S., Egen-McCutchan F., Framson P., Bornstein P.;
RA        "Partial amino acid sequence of human thrombospondin as determined by
RT        analysis of cDNA clones: homology to malarial circumsporozoite
RT        proteins.";
RL        Biochemistry 25:8418-8425(1986).
[4]
RN        SEQUENCE OF 1-374 FROM N.A.
RP        MEDLINE-86287276; PubMed-2461443;
RX        Dixit V.M., Hennessy S.W., Grant G.A., Rotwein P., Frazier W.A.;
RT        "Characterization of a cDNA encoding the heparin and collagen binding
RT        domains of human thrombospondin.";
RL        Proc. Natl. Acad. Sci. U.S.A. 83:5449-5453(1986).
[5]
RN        SEQUENCE OF 1-166 FROM N.A.
RP        MEDLINE-89291870; PubMed-2544587;
RX        Laherty C.D., Glerman T.M., Dixit V.M.;
RT        "Characterization of the promoter region of the human thrombospondin
RT        gene. DNA sequences within the first intron increase transcription.";
RL        J. Biol. Chem. 264:11222-11227(1989).
[6]
RN        SEQUENCE OF 1028-1170 FROM N.A.
RP        la Fleur M., Jobin C., Gauthier J., Kreis C.G.;
RL        Submitted (XXI-1992) to the EMBL/GenBank/DBJ databases.
CC      -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CELL-TO-MATRIX INTERACTIONS CAN BIND TO FIBRINOGEN, FIBRONECTIN,

```

[illegible]

```

FT CARBOHYD 248 360 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 360 360 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 708 708 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1067 1067 T -> A (IN REF. 2, 3 AND 4).
CC CONFLICT 84 84 T -> A (IN REF. 2).
FT CONFLICT 523 523 T -> A (IN REF. 2).
SO SEQUENCE 1170 AA; 129412 MW; 69B3E5DE5AE3A35E CRC64;

Query Match      26.6%; Score 67; DB 1; Length 1170;
Best Local Similarity 39.5%; Pred. No. 1.9;
Matches 15; Conservative 6; Mismatches 11; Indels 6; Gaps 2.

OY 13 ROCMOLE-----TSGOMRRC-VSQGDKRFREDIDMSKY 44
   | | | | | : | | | | | : | | | | | : | | | | |
Db 404 RSCDSLNNKRCGGSVQTRCHTOECDFKRFGQGGSW 441

RESULT 5
TSPL_MOUSE ID TSPL_MOUSE STANDARD; PRT; 1170 AA.
AC P35441;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE THROMBOSPONDIN 1 PRECURSOR.
GN THBS1 OR TSPL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=92128941; PubMed=1774063;
RA Lawler J., Duquette M., Ferro P., Copeland N.G., Gilbert D.J.,
RA Jenkins N.A.;
RT "Characterization of the murine thrombospondin gene.";
RL Genomics 11:587-600(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92147683; PubMed=1371115;
RA Laherty C.D., O'Rourke K., Wolf F.W., Katz R., Seldin M.F.,
RA Dixit V.M.;
RT "Characterization of mouse thrombospondin 2 sequence and expression
during cell growth and development.";
RL J. Biol. Chem. 267:3274-3281(1992).
RN [3]
RP SEQUENCE OF 1-490 FROM N.A.
RX MEDLINE=90375546; PubMed=2398070;
RA Bornstein P., Afifi D., Devarayalu S., Framson P., Li P.;
RT "Characterization of the mouse thrombospondin gene and evaluation of
the role of the first intron in human gene expression.";
RL J. Biol. Chem. 265:16691-16698(1990).
CC -I- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
LAMININ AND TYPE V COLLAGEN.
CC -I- SUBUNIT: HOMOTRIMER, CROSS-LINKED BY DISULFIDE BONDS.
CC -I- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -I- SIMILARITY: CONTAINS 1 VMPC DOMAIN.
CC -I- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -I- SIMILARITY: CONTAINS 3 TYPE-1 TSP REPEATS AND 7 TYPE-3 TSP REPEATS
(WHICH BIND CALCIUM).
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation at
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
RR EMBL; M62470; AAA50611.1; -.
RR EMBL; M62450; AAA50611.1; JOINED.
RR EMBL; M62451; AAA50611.1; JOINED.
```

FT	CARBOHYD	708	708	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1067	1067	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	1025	1025	F -> L (IN AAAS3063).
SO	SEQUENCE	1170 AA;	129646 MW;	0433E93615E7F06 CRC64;

Query Match

Best Local Similarity 39.5%; Pred. No. 1.9;

Matches 15; Conservative 6; Mismatches 11; Indels 6; Gaps 2;

QY 13 RCMQLE-----TSQMRC-VSQCCKFEEDIDMSKY 44

Db 404 RSCDSLNRRCGSSVQFRTCHIQECDRKFKODGMSHW 441

RESULT 6

TSPL_XENLA ID TSPL_XENLA STANDARD; PRT; 1173 AA.

AC P35448;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE THROMBOSPONDIN 1 PRECURSOR.

GN THBS1 OR TSPI.

OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;

OC Xenopodinae; Xenopus.

RN [1]

RP

RA SEQUENCE FROM N.A.

RL Urry L.A., Ramos J., Duquette M., Desimone D.W., Lawler J.;

CC Submitted (xxx-1993) to the EMBL/GenBank/DBJ databases

CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND

CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,

CC LAMININ AND TYPE V COLLAGEN.

CC -1- SUBUNIT: HOMOTRIMER, CROSS-LINKED BY DISULFIDE BONDS.

CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.

CC -1- SIMILARITY: CONTAINS 1 VMWC DOMAIN.

CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.

CC -1- SIMILARITY: CONTAINS 3 TYPE-1 TSP REPEATS AND 7 TYPE-3 TSP REPEATS

CC (WHICH BIND CALCITON).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; L04278; -. NOT_ANNOTATED_CDS.

DR HSSP; P00740; IIXA.

DR INTERPRO; IPR000561; -.

DR INTERPRO; IPR000884; -.

DR INTERPRO; IPR001007; -.

DR PRAM; PR00008; EGF; 2.

DR PRAM; PR00090; TSP_1; 3.

DR PRAM; PR00093; VMWC; 1.

DR PROSITE; PS00022; EGF_1; FALSE_NEG.

DR PROSITE; PS01186; EGF_2; 1.

DR PROSITE; PS50092; TSPI; 3.

DR PROSITE; PS01206; VMWC; 1.

KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;

KW EGF-like domain; Signal.

FT SIGNAL 1 22

FT CHAIN 23 1173

FT DOMAIN 23 235

FT DOMAIN 319 376

FT DOMAIN 382 546

FT DOMAIN 550 693

FT DOMAIN 726 953

FT DOMAIN 954 1173

FT DOMAIN 382 433

REPEAT TSP TYPE-1 1.

POTENTIAL.

THROMBOSPONDIN 1.

HEPARIN-BINDING (POTENTIAL).

VMWC.

3 X TSP TYPE-1 REPEATS (CS-LIKE).

3 X EGF-TYPE REPEATS.

7 X TSP TYPE-3 REPEATS (CA-BINDING).

C-TERMINAL.

TSP TYPE-1 1.

```

FT REPEAT 438 494 TSP TYPE-1 2.
FT REPEAT 495 546 TSP TYPE-1 3.
FT DOMAIN 550 590 EGF-LIKE 1.
FT DOMAIN 591 648 EGF-LIKE 2. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 649 693 EGF-LIKE 3.
FT REPEAT 726 761 TSP TYPE-3 1.
FT REPEAT 762 784 TSP TYPE-3 2.
FT REPEAT 785 820 TSP TYPE-3 3.
FT REPEAT 821 843 TSP TYPE-3 4.
FT REPEAT 844 881 TSP TYPE-3 5.
FT REPEAT 882 917 TSP TYPE-3 6.
FT REPEAT 918 953 TSP TYPE-3 7.
FT SITE 929 931 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 554 565 BY SIMILARITY.
FT DISULFID 559 575 BY SIMILARITY.
FT DISULFID 578 589 BY SIMILARITY.
FT DISULFID 595 611 BY SIMILARITY.
FT DISULFID 602 620 BY SIMILARITY.
FT DISULFID 623 647 BY SIMILARITY.
FT DISULFID 653 666 BY SIMILARITY.
FT DISULFID 660 679 BY SIMILARITY.
FT CARBOHYD 681 692 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 155 155 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 158 158 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 363 363 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 705 705 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 711 711 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1070 1070 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1173 AA: 130019 MW: 49F036D6516C0F24 CRC64;

```

Query Match 25.8%; Score 65; DB 1; Length 1173;
 Best Local Similarity 39.5%; Pred. No. 3.2;
 Matches 15; Conservative 5; Mismatches 12; Indels 6; Gaps 2;

OY 13 RCHQMLE-----TSGQMRRC-VSCDKRFEEDIDMSKY 44
 Db 407 RSCDNLNPGCESSVOTRSCQIDCKRFKODGMSHW 444

```

RESULT 7
SBP_SOYBN STANDARD; PRT; 524 AA.
AC 004672;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE 01-JUN-1994 (Rel. 29, Last annotation update)
DE SUCROSE-BINDING PROTEIN PRECURSOR (SBP).
GN SBP.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Glycine.
RN [1]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 30-50.
RC TISSUE-EMBRYO;
RX MEDLINE=93104680; Pubmed=1467654;
RA Grimes H.D., Overmoore P.J., Ripp K., Franceschi V.R., Hitz W.D.;
RT "A 62-kD sucrose binding protein is expressed and localized in
RT tissues actively engaged in sucrose transport."
RL Plant Cell 4:1561-1574(1992).
CC -1- FUNCTION: PLAYS A ROLE IN SUCROSE TRANSPORT.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED.
CC -1- TISSUE SPECIFICITY: ASSOCIATED WITH THE PLASMA MEMBRANE OF
CC SEVERAL CELL TYPES ENGAGED IN SUCROSE TRANSPORT, INCLUDING THE
CC MESOPHYLL CELLS OF YOUNG SINK LEAVES, THE COMPANION CELLS OF
CC MATURE PHLOEM AND THE CELLS OF DEVELOPING COTYLEDONS.
CC -1- DEVELOPMENTAL STAGE: IN THE COTYLEDON, EXPRESSION IS NOT DETECTED
CC UNTIL 10 DAYS AFTER FERTILIZATION, BETWEEN 10-19 DAYS AFTER
CC FERTILIZATION, EXPRESSION INCREASES RAPIDLY BUT DECLINES 20-30
CC DAYS AFTER FERTILIZATION. 30 DAYS AFTER FERTILIZATION, NO
CC EXPRESSION OCCURS. THIS EXPRESSION PATTERN CLOSELY PARALLELS THE

```

```

CC CC RATE OF SUCROSE UPTAKE IN THE COTYLEDON.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR EMBL: L06038; AAB03894.1; -.
DR PIR: J01730; J01730.
DR HSSP: P50477; ICAW.
DR INTERPRO: IPR00113; -.
DR Pfam: PF00546; Seedstore_7s; 1.
KW Transport; Sugar transport; Signal; Membrane.
FT SIGNAL 1 29
FT CHAIN 30 524 SUCROSE-BINDING PROTEIN.
SQ SEQUENCE 524 AA: 60522 MW: 0251EE90796EF341 CRC64;

```

Query Match 25.0%; Score 63; DB 1; Length 524;
 Best Local Similarity 38.2%; Pred. No. 2.6;
 Matches 13; Conservative 4; Mismatches 15; Indels 2; Gaps 1;

OY 2 EFDROEYECRCHQMLE--TSGQMRRCVSQCDK 33
 Db 34 EEDPELVTCNHCQCCQCCQYTEGDKRVCLQSCDR 67

```

RESULT 8
ATIL_VACCV STANDARD; PRT; 724 AA.
AC P24758;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE 94 KDA A-TYPE INCLUSION PROTEIN (ATI).
OS Vaccinia virus (strain WR).
OC Vaccinia virus (strain WR).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91310644; Pubmed=1856205;
RA Amegadzie B.Y., Ahn B.-Y., Moss B.;
RT "Identification, sequence, and expression of the gene encoding a Mr
RT 35,000 subunit of the vaccinia virus DNA-dependent RNA polymerase."
RL J. Biol. Chem. 266:13712-13718(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92124754; Pubmed=1733111;
RA Amegadzie B.Y., Sisler J.R., Moss B.;
RT "Frame-shift mutations within the vaccinia virus A-type inclusion
RT protein gene."
RL Virology 186:777-782(1992).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92074241; Pubmed=1962448;
RA de Carlos A., Paez E.;
RT "Isolation and characterization of mutants of vaccinia virus with a
RT modified 94-kDa inclusion protein."
RL Virology 185:768-778(1991).
CC -1- FUNCTION: MATURE VIRUSES ARE OCCURED INTO THE ATI, AND IT HAS
CC BEEN ASSUMED THAT SUCH BODIES PROTECT THE VIRUS DURING
CC DISSEMINATION FROM ANIMAL TO ANIMAL.
CC -1- MISCELLANEOUS: A CHARACTERISTIC FEATURE OF ATI IS THE FORMATION OF
CC LARGE MASSES WITH NO SURROUNDING MEMBRANES IN THE CYTOPLASM OF
CC INFECTED CELLS.
CC -1- SIMILARITY: 92% IDENTITY TO COMPOX VIRUS A-TYPE INCLUSION
CC PROTEIN N-TERMINAL.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: M6187: AAA48321.1: -
 DR EMBL: M76371: AAA48275.1: -
 DR EMBL: X57318: CAA40574.1: -
 DR PIR: A41701: MWV294.
 DR PIR: S29908: S29908.
 FT CONFLICT 587 588 KO -> SK (IN REF. 3).
 FT CONFLICT 610 610 R -> H (IN REF. 3).
 FT CONFLICT 618 618 R -> RR (IN REF. 3).
 FT CONFLICT 682 682 S -> R (IN REF. 3).
 SO SEQUENCE 724 AA: 84235 MW: C560CF61C9903028 CRC64;

Query Match 22.6%; Score 57; DB 1; Length 724;
 Best Local Similarity 40.6%; Pred. No. 17;
 Matches 13; Conservative 11; Mismatches 6; Indels 2; Gaps 2;

OY 12 KRCOMLETSQMRRCVSCDCKREEDIDMSK 43
 DB 550 KRRNWEWELS-RLRRDIKECDK-YKEDLDKAK 579

RESULT 9

ATI_CAMPC STANDARD; PRT; 726 AA.

AC 005482;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DE A-TYPE INCLUSION PROTEIN (ATI).
 OS Camelox virus (strain Cp-1).
 OC Viruses: dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Orthopoxvirus.

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93346980; Pubmed=8345359;
 RA Meyer H., Rziha H.-J.;
 RT "Characterization of the gene encoding the A-type inclusion protein
 of camelox virus and sequence comparison with other
 orthopoxviruses.";
 RL J. Gen. Virol. 74:1679-1684(1993).
 CC -1- FUNCTION: MATURE VIRUSES ARE OCCCLUDED INTO THE ATI, AND IT HAS
 BEEN ASSUMED THAT SUCH BODIES PROTECT THE VIRUS DURING
 DISSEMINATION FROM ANIMAL TO ANIMAL.
 CC -1- MISCELLANEOUS: A CHARACTERISTIC FEATURE OF ATI IS THE FORMATION OF
 LARGE MASSES WITH NO SURROUNDING MEMBRANES IN THE CYTOPLASM OF
 INFECTED CELLS.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: X69774: CAA49429.1: -
 DR PIR: JQ2162: JQ2162.
 DR PIR: S32022: S32022.
 KM Late protein: Repeat.
 FT DOMAIN 609 719 4 X APPROXIMATE TANDEM REPEATS.
 FT REPEAT 609 636 1.
 FT REPEAT 637 664 2.
 FT REPEAT 665 688 3.
 FT REPEAT 689 719 4.
 SO SEQUENCE 726 AA: 84415 MW: 1DB20A175C786F20 CRC64;

Query Match 22.6%; Score 57; DB 1; Length 726;
 Best Local Similarity 40.6%; Pred. No. 17;
 Matches 13; Conservative 11; Mismatches 6; Indels 2; Gaps 2;

OY 12 KRCOMLETSQMRRCVSCDCKREEDIDMSK 43
 DB 551 KRRNWEWELS-RLRRDIKECDK-YKEDLDKAK 580

RESULT 10

ATI_COMPX STANDARD; PRT; 1284 AA.

AC P16602;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 01-NOV-1990 (Rel. 16, Last annotation update)
 DE A-TYPE INCLUSION PROTEIN (ATI).
 OS Cowpox virus (CPV).
 OC Viruses: dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Orthopoxvirus.

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CEP05;
 RX MEDLINE=88089536; Pubmed=2826668;
 RA Funahashi S., Sato T., Shida H.;
 RT "Cloning and characterization of the gene encoding the major protein
 of the A-type inclusion body of cowpox virus.";
 RL J. Gen. Virol. 69:35-47(1988).
 CC [2]

CC SEQUENCE OF 1-109 FROM N.A.
 RX MEDLINE=8811568; Pubmed=2828037;
 RA Patel D.D., Pickup D.J.;
 RT "Messenger RNAs of a strongly-expressed late gene of cowpox virus
 contain 5'-terminal poly(A) sequences.";
 RL EMBO J. 6:3787-3794(1987).
 CC -1- FUNCTION: MATURE VIRUSES ARE OCCCLUDED INTO THE ATI, AND IT HAS
 BEEN ASSUMED THAT SUCH BODIES PROTECT THE VIRUS DURING
 DISSEMINATION FROM ANIMAL TO ANIMAL.
 CC -1- MISCELLANEOUS: A CHARACTERISTIC FEATURE OF ATI IS THE FORMATION OF
 LARGE MASSES WITH NO SURROUNDING MEMBRANES IN THE CYTOPLASM OF
 INFECTED CELLS.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: D00319: BAA00222.1: -
 DR EMBL: X06343: CAA29650.1: -
 DR PIR: JQ0006: MWVZAT.
 KM Late protein: Repeat.
 FT DOMAIN 611 912 10 X APPROXIMATE TANDEM REPEATS.
 FT REPEAT 611 637 1.
 FT REPEAT 638 665 2.
 FT REPEAT 666 689 3.
 FT REPEAT 690 720 4.
 FT REPEAT 721 751 5.
 FT REPEAT 752 780 6.
 FT REPEAT 781 811 7.
 FT REPEAT 812 842 8.
 FT REPEAT 843 871 9.
 FT REPEAT 872 912 10.
 SO SEQUENCE 1284 AA: 150329 MW: F7904C9E1DE8D012 CRC64;

Query Match 22.6%; Score 57; DB 1; Length 1284;
 Best Local Similarity 40.6%; Pred. No. 29;
 Matches 13; Conservative 11; Mismatches 6; Indels 2; Gaps 2;

OY 12 KRCOMLETSCOMRCVSOCDKREEDIDMSK 43
 DB 552 KRRVWELSL-RLRRDIKECDK-YKEDIDKAK 581

RESULT 11

PP6C_RAT ID PP6C_RAT STANDARD; PRT; 305 AA.
 AC 064620;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-1998 (Rel. 36, Last annotation update)
 DE SERINE/THREONINE PROTEIN PHOSPHATASE 6 (EC 3.1.3.16) (PP6) (PROTEIN PHOSPHATASE V) (PP-V).
 GN PP6C OR PPV.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=TESTIS, AND BRAIN;
 RX MEDLINE=94357899; PubMed=8077208;
 RA Becker W., Kentrup H., Klump S., Schultz J.E., Joost H.G.:
 RT "Molecular cloning of a protein serine/threonine phosphatase containing a putative regulatory tetrapeptide repeat domain."
 RL J. Biol. Chem. 269:22586-22592(1994).
 CC -1- FUNCTION: MAY FUNCTION IN CELL CYCLE REGULATION.
 CC -1- CATALYTIC ACTIVITY: A PHOSPHOPROTEIN + H(2)O = A PROTEIN + ORTHOPHOSPHATE (THIS ENZYME IS SERINE/THREONINE SPECIFIC).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS FOUND IN SPLEEN, BRAIN AND LUNG.
 CC -1- SIMILARITY: BELONGS TO THE PPP FAMILY OF PHOSPHATASES. PP-V SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X77236; CAA54453.1; -;
 DR HSSP: P08129; IEFM.
 DR INTERPRO: IPR000934; -;
 DR PRAM: PF00149; STPHSPATASE; 1.
 DR PRINTS: PR00114; STPHPTASE.
 DR PROSITE: PS00125; SER_THR_PHOSPHATASE; 1.
 KM Hydrolyase; Iron; Manganese.
 FT METAL 53 53 IRON (BY SIMILARITY).
 FT METAL 55 55 IRON (BY SIMILARITY).
 FT METAL 81 81 IRON AND MANGANESE (BY SIMILARITY).
 FT METAL 113 113 MANGANESE (BY SIMILARITY).
 FT ACT_SITE 114 114 GENERAL ACID (BY SIMILARITY).
 FT METAL 163 163 MANGANESE (BY SIMILARITY).
 FT METAL 237 237 MANGANESE (BY SIMILARITY).
 SO SQUENCE 305 AA; 35106 MW; A91F9939FDBDF110 CRC64;

Query Match 21.8%; Score 55; DB 1; Length 305;
 Best Local Similarity 38.2%; Pred. No. 13;
 Matches 13; Conservative 2; Mismatches 19; Indels 0; Gaps 0;

OY 4 DROEYECRQCMOLETSCOMRCVSOCDKREEF 37
 DB 5 DLDKYVEIAROCKYLPENDLKLCDYVCDLLEE 38

RESULT 12
 ENPL_CHICK ID ENPL_CHICK STANDARD; PRT; 795 AA.

AC P08110; Q90870; Q90869;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE ENDOPLASMIN PRECURSOR (HEAT SHOCK 108 KDA PROTEIN) (HSP108) (HSP 108) (TRANSFERRIN-BINDING PROTEIN).
 GN TRAF.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87076542; PubMed=3024703;
 RA Kilomaa M.S., Weigel N.L., Kleinsek D.A., Beattie W.G., Conneely O.M., March C., Zarucki-Schulz T., Schrader W.T., O'Malley B.W.:
 RT "Amino acid sequence of a chicken heat shock protein derived from the complementary DNA nucleotide sequence."
 RL Biochemistry 25:6244-6251(1986).
 RN [2]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=94220155; PubMed=8166742.
 RA Hayes G.R., Himmler B.S., Weiner K.X.B., Lucas J.J.:
 RT "A chicken transferrin binding protein is heat shock protein 108."
 RL Biochem. Biophys. Res. Commun. 200:65-70(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=OVIDUCT;
 RX MEDLINE=87117509; PubMed=3027654;
 RA Kleinsek D.A., Beattie W.G., Tsai M.J., O'Malley B.W.:
 RT "Molecular cloning of a steroid-regulated 108k heat shock protein gene from hen oviduct."
 RL Nucleic Acids Res. 14:10053-10069(1986).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=OVIDUCT;
 RA Forsgren M.:
 RL Submitted (XXX-1987) to the EMBL/Genbank/DBJ databases.
 CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.
 CC -1- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN.
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M14772; AAA48826.1; -;
 DR EMBL: M31321; AAA48827.1; -;
 DR EMBL: X04961; CAA28629.1; -;
 DR PIR: A24461; HHC08.
 DR PIR: JG2205; JC2205.
 DR HSSP: P02829; JAH8.
 DR INTERPRO: IPR000886; -;
 DR INTERPRO: IPR001404; -;
 DR PRAM: PF00183; HSP90; 1.
 DR PRINTS: PR00775; HEATSHOCK90.
 DR PROSITE: PS00014; ER_TARGET; 1.
 DR PROSITE: PS00298; HSP90; 1.
 KM Chaperone; Endoplasmic reticulum; glycoprotein; calcium binding; K⁺ signal; Heat shock.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 795 ENDOPLASMIN.
 FT CARBOHYD 61 61 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 106 106 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 216 216 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 480 480 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT SITE 792 795 PREVENT SECRETION FROM ER.

```

FT CONFLICT 206 206 D -> E (IN REF. 3 AND 4).
FT CONFLICT 267 267 V -> L (IN REF. 3).
FT CONFLICT 303 303 E -> Q (IN REF. 3).
FT CONFLICT 307 307 N -> D (IN REF. 3 AND 4).
FT CONFLICT 317 317 E -> H (IN REF. 3).
FT CONFLICT 378 378 G -> A (IN REF. 3).
FT CONFLICT 593 594 EG -> DR (IN REF. 3).
FT CONFLICT 653 653 W -> C (IN REF. 3).
FT CONFLICT 669 675 GNDISTN -> VFSS (IN REF. 4).
SQ SEQUENCE 795 AA: 91555 MW: BELB29E1DBEC5A9A CRC64;

```

```

Query Match 21.8%; Score 55; DB 1; Length 795;
Best Local Similarity 23.8%; Pred. No. 32;
Matches 10; Conservative 12; Mismatches 20; Indels 0; Gaps 0;

```

```

OY 2 EPDROEYECRRCOMLETSQMRVCSCDCKRFEEDIDMSK 43
DB 581 EFDGKRFRONAVAKGVKFEESKSKSEHLEKEFEPLLMNMK 622

```

```

RESULT 13
YL53_CAEEL STANDARD; PRT; 244 AA.
AC P34453;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE HYPOTHETICAL 29.0 KDA PROTEIN F44E2.3 IN CHROMOSOME III.
GN F44E2.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
NC [1]
RC STRAIN-BRISTOL N2;
RX MEDLINE-94150718; PubMed-7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latelle P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smalldon N., Smith A., Smith M., Soudhamer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaughan K.,
RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Woldman P.;
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.
RL Nature 368:32-38(1994).
CC -1- SIMILARITY: TO DNABJ.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL: L23646; AAA28041.1; -
DR PIR: S44822; S44822.
DR MORPEP: F44E2.3; CE00181.
KM Hypothetical protein.
FT DOMAIN 3 45 ARG/ASP-LYS-RICH.
FT 79 90 PRO-RICH.
SQ SEQUENCE 244 AA: 28994 MW: C0CE677FB01A2B18 CRC64;

```

```

Query Match 21.6%; Score 54.5; DB 1; Length 244;
Best Local Similarity 39.4%; Pred. No. 12;
Matches 13; Conservative 8; Mismatches 7; Indels 5; Gaps 2;

```

```

OY 9 EECRRCOMLETSQMRVCSCDCKRFEEDIDMSK 38
DB 164 EESRKKCKOLE-ALERKVLVEAESRKKFEED 194

```

```

RESULT 14
YL68_CAEEL STANDARD; PRT; 1021 AA.
AC P46582;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE HYPOTHETICAL 111.9 KDA PROTEIN C34E10.8 IN CHROMOSOME III.
GN C34E10.8
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
NC [1]
RC STRAIN-BRISTOL N2;
RX MEDLINE-94150718; PubMed-7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latelle P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smalldon N., Smith A., Smith M., Soudhamer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaughan K.,
RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Woldman P.;
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.
RL Nature 368:32-38(1994).
CC -1- SIMILARITY: TO DNABJ.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL: U10402; AAA19070.1; -
DR MORPEP: C34E10.8; CE01188.
KM Hypothetical protein.
SQ SEQUENCE 1021 AA: 111858 MW: F1B63DAA26332F5F CRC64;

```

```

Query Match 21.6%; Score 54.5; DB 1; Length 1021;
Best Local Similarity 22.0%; Pred. No. 46;
Matches 11; Conservative 15; Mismatches 13; Indels 11; Gaps 2;

```

```

OY 1 SEPDROEYECRRCOMLETSQMRVCSCDCKRFEEDIDMSK 43
DB 24 NEYDKNRYGOCNAYSYRTLVRCQIRSKELARHGRC---BEHVEFSK 69

```

```

RESULT 15
CGD1_BRARE STANDARD; PRT; 291 AA.
AC C90459;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE G1/S-SPECIFIC CYCLIN D1.
GN CYCLD.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
NC [1]
RC STRAIN-BRISTOL N2;
RX MEDLINE-96138542; PubMed-8547308;
RA Yarden A., Salomon D., Geiger B.;
RT *zebrafish cyclin D1 is differentially expressed during early
RT embryogenesis.
RL Biochim. Biophys. Acta 1264:257-260(1995).
CC -1- FUNCTION: ESSENTIAL FOR THE CONTROL OF THE CELL CYCLE AT THE G1/S
CC (START) TRANSITION.
CC -1- SUBUNIT: INTERACTS WITH THE CDK4 AND CDK6 PROTEIN KINASES TO FORM
CC A SERINE/THREONINE KINASE HOLOENZYME COMPLEX. THE CYCLIN SUBUNIT
CC IMPARTS SUBSTRATE SPECIFICITY TO THE COMPLEX.
CC -1- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN D SUBFAMILY.

```


